Geophysical Research Abstracts
Vol. 15, EGU2013-11531, 2013
EGU General Assembly 2013

# A spatio-temporal screening tool for outlier detection in long term / large scale air quality observation time series and monitoring networks

Oliver Kracht, Hannes I. Reuter, and Michel Gerboles

European Commission - Joint Research Centre, Institute for Environment and Sustainability, 21020 Ispra (VA), Italy (oliver.kracht@jrc.ec.europa.eu)

We present a consolidated screening tool for the detection of outliers in air quality monitoring data, which considers both attribute values and spatio-temporal relationships. Furthermore, an application example of warnings on abnormal values in time series of PM10 datasets in AirBase is presented.

Spatial or temporal outliers in air quality datasets represent stations or individual measurements which differ significantly from other recordings within their spatio-temporal neighbourhood. Such abnormal values can be identified as being extreme compared to their neighbours, even though they do not necessarily require to differ significantly from the statistical distribution of the entire population. The identification of such outliers can be of interest as the basis of data quality control systems when several contributors report their measurements to the collection of larger datasets. Beyond this, it can also provide a simple solution to investigate the accuracy of station classifications. Seen from another viewpoint, it can be used as a tool to detect irregular air pollution emission events (e.g. the influence of fires, wind erosion events, or other accidental situations).

The presented procedure for outlier detection was designed based on already existing literature. Specifically, we adapted the “Smooth Spatial Attribute Method” that was first developed for the identification of outlier values in networks of traffic sensors [1]. Since a free and extensible simulation platform was considered important, all codes were prototyped in the R environment which is available under the GNU General Public License [2].

Our algorithms are based on the definition of a neighbourhood for each air quality measurement, corresponding to a spatio-temporal domain limited by time (e.g., +/- 2 days) and distance (e.g., +/- 1 spherical degrees) around the location of ambient air monitoring stations. The objective of the method is that within such a given spatio-temporal domain, in which the attribute values of neighbours have a relationship due to the emission, transport and reaction of air pollutants, abnormal values can be detected by extreme values of their attributes compared to the attribute values of their neighbours. This comparison basically requires a spatio-temporal smoothing, i.e. a specific rule by which data points are averaged within a neighbourhood. The calculation of such reference basis has the effect of a low pass filter, meaning that high frequencies of the signal are removed from the data while preserving low frequencies. In this context, the choice of an appropriate kernel smoother function (e.g., nearest neighbour smoother, weighted kernel average smoother, etc.) is of particular importance. Our presentation will emphasize the effects bound to the selection of the corresponding weighting functions, like inverse squared normalized Euclidean distance or inverse squared Mahalanobis distance etc., and discuss the appropriateness and shortcomings of the different approaches. Corresponding parameter selections related to the extent of the spatio-temporal domain and the final test statistics for outlier thresholding are evaluated by sensitivity analysis.

[1] R Development Core Team (2012): R: A language and environment for statistical computing. http://www.R-project.org/

[2] Shekhar S., Lu C. T. and Zhang P. (2003): A Unified Approach to Spatial Outliers Detection. GeoInformatica, 7(2).